Balance conditions in variational data assimilation for a high-resolution forecast model

Received: 21 October 2020 | Revised: 1 May 2021 | Accepted: 5 May 2021 | Published on: 9 June 2021

DOI: 10.1002/qj.4106

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Balance conditions in variational data assimilation for a high-resolution forecast model

Ross N. Bannister

Department of Meteorology and National Centre for Earth Observation, University of Reading, Reading, UK

Correspondence
R.N. Bannister, Department of Meteorology, University of Reading, Whiteknights Road, Earley Gate, Reading RG6 6ET, UK.
Email: r.n.bannister@reading.ac.uk

Funding information
U.K. Natural Environment Research Council via the National Centre for Earth Observation, Grant/Award Number: LTS-S NE/R016518/1

Abstract

This paper explores the role of balance relationships for background-error covariance modelling as the model's grid box decreases to convective scales. Data assimilation (DA) analyses are examined from a simplified convective-scale model and DA system (called ABC-DA) with a grid box size of 1.5 km in a 2D 540 km (longitude), 15 km (height) domain. The DA experiments are performed with background-error covariance matrices (**B**) modelled and calibrated by switching on/off linear balance (LB) and hydrostatic balance (HB), and by observing a subset of the ABC variables, namely v (meridional wind), $\widetilde{\rho}'$ (scaled density, a pressure-like variable), and b' (buoyancy, a temperature-like variable). Calibration data are sourced from two methods of generating proxies of forecast errors. One uses forecasts from different latitude slices of a 3D parent model (here called the latitude slice method), and the other uses sets of differences between forecasts of different lengths but valid at the same time (the National Meteorological Center method). Root-mean-squared errors computed over the domain from identical twin DA experiments suggest that there is no combination of LB/HB switches that give the best analysis for all model quantities. However it is frequently found that the **B**-matrices modelled with both LB and HB do perform the best. A clearer picture emerges when the errors are examined at different spatial scales. In particular it is shown that switching on HB in **B** mostly has a neutral/positive effect on the DA accuracy at 'large' scales, and switching off the HB has a neutral/positive effect at 'small' scales. The division between 'large' and 'small' scales is between 10 and 100 km. Furthermore, one hour forecast-error correlations computed between control parameters find that correlations are small at large scales when balances are enforced, and at small scales when balances are not enforced (ideal control parameters have zero cross-correlations). This points the way to modelling **B** with scale-dependent balances.

KEYWORDS

balance, convective scale, data assimilation

1 | INTRODUCTION

Balance constraints of one form or another have been used to help formulate data assimilation (DA) problems for many years, e.g., Lorenc (1981); Derber and Bouttier (1999); Gao *et al.* (1999); Berre (2000); Ingleby (2001); Fisher (2003); Ge *et al.* (2012); Tong *et al.* (2016). Balance constraints are useful for a number of reasons, especially (though not exclusively) when used with variational and hybrid schemes. Firstly, they are used to dampen damaging unbalanced motion produced spuriously as a by-product of the analysis procedure. This is done by building them into the formulation of the static background-error covariance matrix, **B** (Parrish and Derber, 1992; Derber and Bouttier, 1999; Gauthier *et al.*, 1999; Bannister, 2008), enforcing them as an extra constraint in the cost function (a J_c term) (Hu *et al.*, 2006; Kleist *et al.*, 2009), and applying them post analysis explicitly or implicitly (Courtier and Talagrand, 1990; Lynch and Huang, 1992; Bloom *et al.*, 1996; Potvin and Wicker, 2013). Secondly, when a balance constraint is imposed strongly, it reduces the number of degrees of freedom that a DA problem has to deal with. Thirdly, and specifically for variational and hybrid DA schemes, balance conditions guide the definition of the control variables, which are assumed to be mutually uncorrelated, thus defining a model of **B** (Bannister, 2008; Song and Kang, 2019).

The choice of balance conditions (if any) relevant to a particular DA problem depends upon the properties of the flow, which in turn depend on the fluid's geographical location and on the scales of motion considered. For instance, geostrophic or hydrostatic balances – dominant in flows of small Rossby numbers typical of the large-scale extratropical free troposphere – are not always well suited to regimes that have a potentially high Rossby number – for example, tropical or small-scale flows (Sun, 2005; Sun *et al.*, 2014; Yano *et al.*, 2018), rain (Caron and Fillion, 2010), convection (Vetra-Carvalho *et al.*, 2012), or high flow curvature (Fisher, 2003). Despite these concerns, balance relationships like the linear and nonlinear balance equations, and hydrostatic balance, whether in dynamical or regressed forms, are associated with the formulation of **B** for km-scale DA systems, for example, Barker *et al.* (2004); Honda *et al.* (2005); Brousseau *et al.* (2011); Ballard *et al.* (2016); Heng *et al.* (2020); Xu *et al.* (2020). As models have smaller and smaller grid lengths, this leads to obvious questions on whether it is an optimal, efficient, or appropriate use of balance conditions in such small-scale systems, and whether they can be dropped or replaced with alternative balance conditions which are more appropriate at convective scales. To date, there have been few systematic studies of 'turning off' balance conditions in **B** of variational or hybrid DA systems. There have been a number of studies that appear to have compared the performance of 3D-Var systems through a change of variational control variables from a set that exploits balance operators to a set that does not (Xie and MacDonald, 2012; Zakeri *et al.*, 2018; Shen *et al.*, 2019; Thiruvengadam *et al.*, 2019; Wang *et al.*, 2020). These studies did find benefits of the change but the change had been made at the same time as a change of momentum control variables from streamfunction and velocity potential-based variables to zonal and meridional wind-based variables. It remains an open question on which balance conditions, if any, remain appropriate for use with convective-scale DA.

This paper revisits the problem of the use of geophysical balance to model **B** as used in variational and hybrid schemes when applied to midlatitude flows that contain small scales that are expected to have a considerable amount of unbalanced flow by their nature. The model used for most of this study is a high-resolution (1.5 km grid length) version of the simplified ABC model (Petrie *et al.*, 2017), and its variational DA system, ABC-DA (Bannister, 2020), but some results use ensemble forecasts from a 1.5 km grid length version of the Met Office's Unified Model (UM) for a limited-area domain over the southern UK.

The structure of this paper is as follows. In Section 2 we briefly describe the ABC-DA system. In Section 3 we summarise the balance relationships considered in this study. In Section 4 we review how they are used in ABC-DA. In Section 5 we show how two populations of forecast-error proxies have been generated, which are used to calibrate our **B** matrices. In Section 6 we discuss the potential problems resulting from inappropriate balance conditions in **B** and demonstrate that such problems are evident in some sample data from the convection-permitting UM and ABC systems. In Section 7 we define and show the ABC-DA experiments performed with different balance conditions applied to **B**. In Section 8 we relate these results to the reasoning made earlier in the paper. Finally in Section 9 we summarise our findings, discuss the possible limitations of this study, and outline further work that could be done. In addition, The Appendix summarises the balance conditions for the Euler equations. Even though our focus (for simplicity) is the ABC-DA system, the conclusions of this paper will be of interest to operational centres, especially those that use 3D/4D-Var and hybrid methods.

2 | THE ABC-DA SYSTEM

The ABC model (Petrie *et al.*, 2017) is a set of simplified fluid equations which were designed to permit speedy research into convective-scale DA. They are modified versions of the compressible Euler equations, designed to

exhibit balanced motion at large scales, but unbalanced motion on smaller scales. The equations operate on a 2D longitude/height plane (x/z); they are partly linearised (advection terms and the mass continuity equations remain nonlinear); the Brunt–Väisälä frequency is taken to be a constant, A (to control gravity wave frequencies), the advection terms and the mass divergence terms are modulated by a parameter $0 < B \leq 1$ (to control acoustic wave speeds), and the equation of state is simplified to relate pressure and density perturbations via an inverse compressibility coefficient, C. The result is a set of five prognostic equations for zonal wind, u, meridional wind, v, vertical wind, w, scaled density perturbation, $\widetilde{\rho}' = \rho'/\rho_0$ (where ρ_0 is reference density), and buoyancy perturbation, $b' = (g/\theta_R)\theta'$ (where g is the acceleration due to gravity, θ_R is the reference potential temperature, and θ' is the potential temperature perturbation). The equations are given as equation (15) of Petrie *et al.* (2017). Additionally, pressure increments may be diagnosed from the simplified equation of state, $p' = C\rho'$. The equations are dry, but work is currently under way to include simplified moist processes. A is the pure gravity wave frequency and $\sqrt{BC}$ is the pure small-scale acoustic wave speed. The parameter B was introduced to slow the speed of the acoustic waves so that the equations can be solved using an explicit scheme (the split-explicit forward-backward scheme; Cullen and Davies 1991). In this paper, the parameters are set as follows: $A = 0.01\,\mathrm{s}^{-1}$, $B = 0.01$, $C = 10^5\,\mathrm{m}^2\mathrm{s}^{-2}$, and the Coriolis parameter is $f = 10^{-4}\mathrm{s}^{-1}$. The domain of the model is 540 km by 15 km and the grid is chosen to have 360 (1.5 km length) horizontal grid cells and 60 vertical levels, which is thought to be on the edge of the 'grey zone' where convective parametrizations are still needed in operational models (Yu and Lee, 2010) (although the ABC model presently captures only dry circulations).

The model is thought to behave in a qualitatively similar way to the real atmosphere. Flows may be approximately decomposed into slow-moving/low-frequency balanced modes, intermediate frequency gravity modes, and high-frequency acoustic modes. A linear analysis of this model performed about a state of rest (not shown) reveals a zero frequency Rossby-like mode, which is in a state of geostrophic and hydrostatic balance, gravity modes with frequencies up to $0.01\,\mathrm{s}^{-1}$ (the value of A) and speeds approaching $20\,\mathrm{ms}^{-1}$, and acoustic modes with frequencies up to $\sim 0.7\,\mathrm{s}^{-1}$ and speeds $\sim 30\,\mathrm{ms}^{-1}$. The fast gravity and acoustic waves allow the model to adjust as a result of added perturbations, for example, due to DA. These aspects, including a demonstration of geostrophic adjustment in ABC, are discussed in Petrie *et al.* (2017).

The associated DA (Bannister, 2020) is an incremental 3D-Var/3DFGAT-based variational system. The **B**-matrix is modelled with a set of parameter/horizontal/vertical transforms where the model variables are decomposed into balanced and unbalanced components which are assumed to be uncorrelated. The ABC-DA control parameters are streamfunction ($\delta\psi$), velocity potential ($\delta\chi_{\mathrm{vp}}$), geostrophically unbalanced scaled density ($\delta\widetilde{\rho}'^{\mathrm{u}}$, akin to unbalanced pressure in other systems), hydrostatically unbalanced buoyancy ($\delta b'^{\mathrm{u}}$, akin to unbalanced temperature, which is not considered in other systems, for example, the Met Office's variational system, which assumes that all temperature increments are hydrostatically balanced), and vertical wind (δw, which is diagnosed from other variables in the Met Office's system). The scheme is flexible to allow the balance conditions used in the **B**-matrix to be switched on or off to study their effects on the analysis. Observations can be of any of the variables at arbitrary positions in space and time. The system includes a suite to calibrate the **B**-matrix from ensembles of possible background states, and a suite to run the assimilation in a cycled forecast/DA mode.

3 | SUMMARY OF BALANCE RELATIONSHIPS FOR THE ABC SYSTEM

In this section, we summarise the balance relationships considered in this paper. Interest is ultimately concerned with the equation sets of operational systems, so we give two versions of the balance relationships: the balance relationships used for the ABC system are given here and those relevant to the Euler equations are given in the Appendix.

3.1 | Linear balance in ABC

The linear balance equation (LBE) in ABC emerges from a scale analysis of the zonal momentum equation for small Rossby number (section 2.2 of Petrie *et al.* (2017)):

$$\delta\widetilde{\rho}'^{\mathrm{b}} = (f/C)\delta\psi, \tag{1}$$

where a δ prefix indicates an increment, and a b superscript indicates the balanced part of the variable[1]. Since f is constant in ABC, Equation (1) is equivalent to geostrophic balance.

[1]Note that in the ABC system, where there is no latitude dependence, the Helmholz relations lead to $\delta v = \partial\delta\psi/\delta x$ and $\delta u = \partial\delta\chi_{\mathrm{vp}}/\partial x$, where $\delta\psi$ is the streamfunction increment and $\delta\chi_{\mathrm{vp}}$ is the velocity potential increment.

3.2 | Hydrostatic balance in ABC

The hydrostatic balance equation (HBE) emerges from a scale analysis of the vertical momentum equation for either small Rossby number or for a small ratio of vertical to horizontal wind magnitude:

$$C\partial\delta\widetilde{\rho}'/\partial z = -\delta b'^{\mathrm{b}}. \tag{2}$$

A strongly imposed hydrostatic balance eliminates vertically propagating acoustic waves. In ABC-DA though, and unlike other systems, some hydrostatically unbalanced analysis increments are allowed (Bannister, 2020).

4 | USING THE BALANCE RELATIONS IN ABC-DA

The ABC-DA variational scheme is documented in Bannister (2020), but the relevant parts of the control variable transform (**U**), which define the **B**-matrix, are summarised as follows. As mentioned in Section 2, the *control parameters* of ABC-DA are $\delta\psi$, $\delta\chi_{\mathrm{vp}}$, $\delta\widetilde{\rho}'^{\mathrm{u}}$, $\delta b'^{\mathrm{u}}$, and δw. These parameters are assumed to be mutually uncorrelated (in the sense of background errors) for the purposes of modelling **B**. These parameters are fields that still have spatial covariances, but are related to the associated *control variables* ($\delta\chi_{\psi}$, $\delta\chi_{\chi_{\mathrm{vp}}}$, $\delta\chi_{\widetilde{\rho}'^{\mathrm{u}}}$, $\delta\chi_{b'^{\mathrm{u}}}$, and $\delta\chi_w$ respectively) via the spatial transform, $\mathbf{U}^{\mathrm{s}}$:

$$\delta\psi = \mathbf{U}^{\mathrm{s}}_{\psi}\delta\chi_{\psi}, \quad \delta\chi_{\mathrm{vp}} = \mathbf{U}^{\mathrm{s}}_{\chi_{\mathrm{vp}}}\delta\chi_{\chi_{\mathrm{vp}}}, \quad \delta\widetilde{\rho}'^{\mathrm{u}} = \mathbf{U}^{\mathrm{s}}_{\widetilde{\rho}'^{\mathrm{u}}}\delta\chi_{\widetilde{\rho}'^{\mathrm{u}}},$$
$$\delta b'^{\mathrm{u}} = \mathbf{U}^{\mathrm{s}}_{b'^{\mathrm{u}}}\delta\chi_{b'^{\mathrm{u}}}, \text{ and } \delta w = \mathbf{U}^{\mathrm{s}}_{w}\delta\chi_{w}.$$

Control variables are taken to have no auto-covariances and to have unit background variances. The $\mathbf{U}^{\mathrm{s}}_{\psi}$, etc., form the block-diagonals of $\mathbf{U}^{\mathrm{s}}$, and are the square-roots of the background-error auto-covariances; for example, $\mathbf{U}^{\mathrm{s}}_{\psi}\mathbf{U}^{\mathrm{s}\mathrm{T}}_{\psi}$ is the auto-covariance matrix of $\delta\psi$. Control variable transforms are described more fully in Bannister (2008).

The key parts of the transforms relevant to this paper are as follows (see also Table 1). The total scaled density increment, $\delta\widetilde{\rho}'$, is found as the sum of balanced, $\delta\widetilde{\rho}'^{\mathrm{b}}$, and unbalanced, $\delta\widetilde{\rho}'^{\mathrm{u}}$, parts where the balanced part is found from $\delta\psi$ via the LBE (Equation (1)) (steps II and III of Table 1):

$$\begin{aligned}\delta\widetilde{\rho}' &= \delta\widetilde{\rho}'^{\mathrm{b}} + \delta\widetilde{\rho}'^{\mathrm{u}} = \alpha(f/C)\delta\psi + \delta\widetilde{\rho}'^{\mathrm{u}}\\ &= \alpha(f/C)\mathbf{U}^{\mathrm{s}}_{\psi}\delta\chi_{\psi} + \mathbf{U}^{\mathrm{s}}_{\widetilde{\rho}'^{\mathrm{u}}}\delta\chi_{\widetilde{\rho}'^{\mathrm{u}}}.\end{aligned} \tag{3}$$

The factor α is introduced to turn on ($\alpha = 1$) or turn off ($\alpha = 0$) the effect of the LBE in the covariance model. The total buoyancy increment, $\delta b'$, is similarly found as the sum of balanced, $\delta b'^{\mathrm{b}}$, and unbalanced, $\delta b'^{\mathrm{u}}$, parts where the balanced part is found from $\delta\widetilde{\rho}'$ via the HBE (Equation (2)) (steps IV and V of Table 1):

$$\begin{aligned}\delta b' &= \delta b'^{\mathrm{b}} + \delta b'^{\mathrm{u}} = -\beta C\frac{\partial}{\partial z}\delta\widetilde{\rho}' + \delta b'^{\mathrm{u}}\\ &= -\beta C\frac{\partial}{\partial z}\left\{\alpha(f/C)\mathbf{U}^{\mathrm{s}}_{\psi}\delta\chi_{\psi} + \mathbf{U}^{\mathrm{s}}_{\widetilde{\rho}'^{\mathrm{u}}}\delta\chi_{\widetilde{\rho}'^{\mathrm{u}}}\right\}\\ &+ \mathbf{U}^{\mathrm{s}}_{b'^{\mathrm{u}}}\delta\chi_{b'^{\mathrm{u}}}.\end{aligned} \tag{4}$$

The factor β is introduced to turn on/off the effect of hydrostatic balance. Note that only $\alpha = 0, 1$ and $\beta = 0, 1$ are considered in detail in this paper, although in the summary we consider treating α and β as continuous variables as a possible extension. These steps are similar to the Met Office's control variable transform (Lorenc *et al.*, 2000; Ingleby, 2001), except that the Met Office (a) has an extra vertical regression step after application of the LBE, and (b) does not allow an unbalanced temperature increment (equivalent to always setting $\delta b'^{\mathrm{u}} = 0$ in ABC). The vertical regression step acts to complete the prediction of the 'balanced' pressure and can allow for inapplicability of enforcing the LBE directly. The vertical regression step is an optional part of ABC-DA, but is not applied here because the earlier study of Bannister (2020) found it to degrade the accuracy of the ABC analysis, and interest in this paper is the application of purely analytical balance relationships.

The assumption that the control variables are uncorrelated leads to the implied **B**-matrix of this system. For instance, when $\alpha = 1$ the implied background-error variance for scaled density emerges from Equation (3) as follows:

$$\begin{aligned}\left\langle\delta\widetilde{\rho}'^{2}\right\rangle = &\underbrace{\left(\frac{f}{C}\right)^{2}\left\langle\delta\psi^{2}\right\rangle}_{\substack{\text{balanced scaled}\\ \text{density variance}}} + \underbrace{\left\langle\delta\widetilde{\rho}'^{\mathrm{u}2}\right\rangle}_{\substack{\text{unbalanced scaled}\\ \text{density variance}}}\\ &+ \underbrace{2\frac{f}{C}\left\langle\delta\psi\delta\widetilde{\rho}'^{\mathrm{u}}\right\rangle}_{\substack{\text{bal./unbal. scaled density}\\ \text{covariance, zero by assumption}}},\end{aligned} \tag{5}$$

TABLE 1 Summary of the parameter transform of the convective-scale scheme detailed in Bannister (2020). Control parameters are underlined, and α and β are switches described in the text

I. Winds	$\delta u = \nabla_x\underline{\delta\chi_{\mathrm{vp}}}$, $\delta v = \nabla_x\underline{\delta\psi}$
II. Balanced mass	$\delta\widetilde{\rho}'^{\mathrm{b}}$ (Equation (1))
III. Total mass	$\delta\widetilde{\rho}' = \alpha\delta\widetilde{\rho}'^{\mathrm{b}} + \underline{\delta\widetilde{\rho}'^{\mathrm{u}}}$
IV. Balanced buoyancy	$\delta b'^{\mathrm{b}}$ (Equation (2))
V. Total buoyancy	$\delta b' = \beta\delta b'^{\mathrm{b}} + \underline{\delta b'^{\mathrm{u}}}$
VI. Vertical wind	$\underline{\delta w}$

where $\langle \bullet \rangle$ is the expectation over the hypothetical background probability density function[2]. The last term in Equation (5) is zero if the zero-correlation assumption of the covariance model is correct. Since this is assumed in the DA, the implied covariance comprises only the first two terms of Equation (5). If exploration of the background-error statistics outside of the DA environment show that this term is non-zero, then this term represents the covariance model's anomaly, and will represent a sub-optimality in the DA. A similar argument applies to b'. When $\alpha = 0$ the balanced variance is not present and so the implied variance is purely unbalanced $\langle \delta\widetilde{\rho}'^2 \rangle = \langle \delta\widetilde{\rho}'^{u^2} \rangle$ and the implied covariance between $\delta\psi$ and $\delta\widetilde{\rho}'$ is lost.

5 | ORIGIN OF THE TWO CALIBRATION ENSEMBLES

Forecast-error statistics are affected by the observation network, the DA system, the length of the prior forecasts, and the errors in the formulation of the model (e.g., Houtekamer *et al.*, 1996), including representativity error (Hodyss and Nichols, 2015) (Model and representativity errors are absent in this study as we perform only identical twin experiments to test the DA.) In this study we follow the standard procedure of attempting to calibrate **B** using a sample from these statistics. However, the generation of such a sample of forecast errors is difficult to do from scratch since it requires the **B**-matrix to define the DA in the first place (to generate the forecasts' initial conditions). This is a 'chicken-and-egg' problem. For this reason, we accept that there is no easy or perfect solution and we proceed pragmatically. We propose two methods of defining populations of forecast errors used to calibrate **B**, and compare how each affects the performance of the DA system. Each method is described below.

5.1 | The latitude slice method

The first method takes sequences of u and v fields from different latitudes of a dump of the Unified Model with the same horizontal grid length and domain size. Each slice is modified so that it obeys periodic boundary conditions (section 5.2 of Petrie *et al.*, 2017). The w field is found by then imposing zero three-dimensional divergence. Since $\widetilde{\rho}'$ and b' are not part of the UM, these are found from the balance conditions (first $\widetilde{\rho}'$ is found by integrating the LBE given v, and then b' is found from the HBE given $\widetilde{\rho}'$). This processing from each latitude provides a set of ABC initial conditions, which are then run through the ABC model for 1 hr. The resulting population of states forms an ensemble of 260 plausible model forecasts, and the deviations from the mean are assumed proxies for forecast errors. We believe that the 1 hr integration is long enough to spin up/down imbalances in the system as revealed by studying variance spectra of balanced and unbalanced $\widetilde{\rho}'$ at a function of lead time. These converge quickly to the 1 hr spectra shown in the next section.

5.2 | The NMC method

The second method derives proxy forecast errors using the standard National Meteorological Centre (NMC) method (Parrish and Derber, 1992) by taking a population of 260 2 hr minus 1 hr forecasts (divided by $\sqrt{2}$) from the cycled ABC-DA system. As this method relies on an existing DA system (including the **B**-matrix), we are faced with a pragmatic choice and choose this underlying system to be that calibrated with statistics from the latitude slice method where all balances in the **B**-matrix are switched off. This is an arbitrary decision, but investigating the effect of the underlying DA system on the NMC method is outside the scope of this work, but for now does provide an alternative set of ABC-DA results to study.

Even though the NMC method is widely used, it does not yield perfect proxies of forecast errors (Berre *et al.*, 2006). For instance, as the method uses differences between forecasts valid at the same time, it is likely to underestimate errors in unobserved quantities (in this study u and w are unobserved; Section 7).

6 | POSSIBLE CONSEQUENCES OF USING BALANCE CONDITIONS INAPPROPRIATELY

The model of **B** that gives Equation (5) is useful only if $\delta\widetilde{\rho}'^{b}$ as found from $\delta\psi$ (Equation (1)) has a strong correlation with $\delta\widetilde{\rho}'$, and the residual, $\delta\widetilde{\rho}'^{u}$, is uncorrelated with $\delta\widetilde{\rho}'^{b}$. This may happen by $\delta\widetilde{\rho}'^{b}$ and $\delta\widetilde{\rho}'^{u}$ having very different time-scales (i.e., by the gravity modes being associated with faster processes than Rossby modes as discussed in Section 2). This is the situation in the schema shown in Figure 1a, which is the case expected for large-scale, midlatitude dynamics where balanced increments may be thought to define a balanced manifold where unbalanced increments form a 'fuzzy' region around that manifold. Alternatively when $\delta\widetilde{\rho}'^{b}$ as found from $\delta\psi$ does not predict

[2]Interested readers are referred to section 4.4 of Bannister (2020) for the full multivariate implied **B**-matrix (where cross-correlations between control variables are assumed zero). Additionally the covariance structures are revealed graphically in single-observation experiments shown in section 7.2 of that paper.

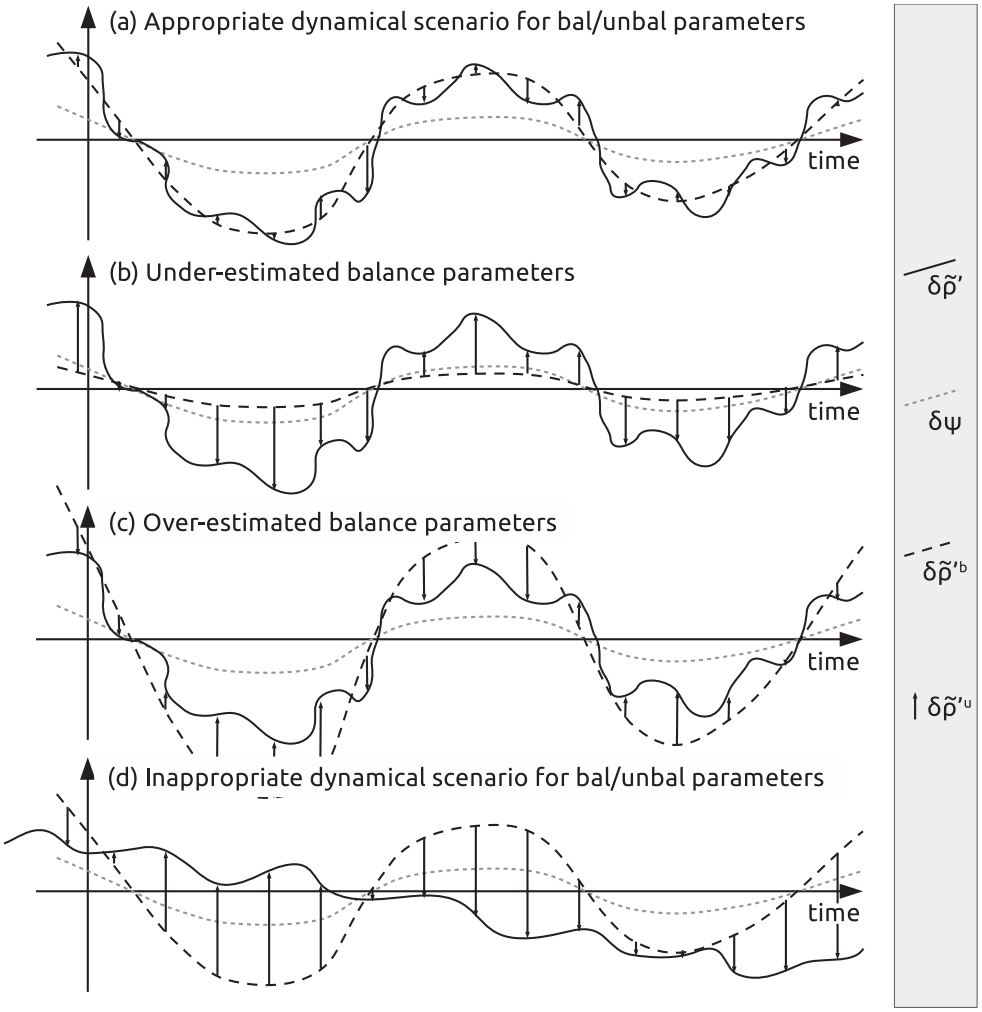

FIGURE 1 Schemas to illustrate situations when (a) the balanced ($\delta\widetilde{\rho}'^{\mathrm{b}}$)/unbalanced ($\delta\widetilde{\rho}'^{\mathrm{u}}$) partitioning of a background-error variable ($\delta\widetilde{\rho}'$) is appropriate for the model of **B**, (b) when the balanced parameter is systematically underestimated by the balance operator, (c) when the balanced parameter is systematically overestimated, and (d) when the balanced/unbalanced partitioning is completely inappropriate. In all panels, $\delta\widetilde{\rho}' = \delta\widetilde{\rho}'^{\mathrm{b}} + \delta\widetilde{\rho}'^{\mathrm{u}}$. In (a) $\delta\widetilde{\rho}'^{\mathrm{b}}$ broadly follows $\delta\widetilde{\rho}'$, and $\delta\widetilde{\rho}'^{\mathrm{u}}$ is mutually uncorrelated with $\delta\widetilde{\rho}'^{\mathrm{b}}$. In (b/c) $\delta\widetilde{\rho}'^{\mathrm{b}}$ and $\delta\widetilde{\rho}'^{\mathrm{u}}$ are expected to be positively/negatively correlated. In (d) this correlation is non-zero of undetermined sign. In cases (b, c, d) the variance of $\delta\widetilde{\rho}'^{\mathrm{u}}$ is expected to be too large when the variational DA statistics are calibrated

well $\delta\widetilde{\rho}'$, there will be a large $\delta\widetilde{\rho}'^{\mathrm{u}}$ to compensate for the bad $\delta\widetilde{\rho}'^{\mathrm{b}}$ prediction. Consequently $\delta\widetilde{\rho}'^{\mathrm{u}}$ and $\delta\psi$ will become correlated (even though they are still *assumed* to be uncorrelated in the covariance model). This is the situation in Figure 1b–d. In cases (b, c), $\delta\widetilde{\rho}'^{\mathrm{b}}$ consistently under/overestimates $\delta\widetilde{\rho}'$ (representing cases when the balance relation is applied not strongly enough/ too strongly) and in case (d) $\delta\widetilde{\rho}'^{\mathrm{b}}$ is completely unrelated to $\delta\widetilde{\rho}'$ (representing cases when there is no useful information in the applied balance relation). The implied variance for $\delta\widetilde{\rho}'$ is the sum of the balanced and unbalanced contributions (only the first two terms in Equation (5)), which may be an under- or overestimate of the true variance, meaning that observations of scaled density will be under- or overfitted by the DA, respectively. The quantity that demonstrates that a covariance model is not appropriate is the last term in Equation (5) (the anomaly), $2\left\langle \widetilde{\rho}'^{\mathrm{b}}\widetilde{\rho}'^{\mathrm{u}}\right\rangle$.

For illustration, Figure 2 shows the contributions to the mass variances at about 4 km elevation as a function of scale as found from ensembles of the UM (a, b at two different times 1230 and 1630 UTC on 20 September 2011) and ABC forecasts (c, d found from the latitude slice and NMC methods; Section 5). The mass variable is pressure and scaled density in the UM and ABC systems respectively, the LBEs Equations (A2) and (1) are used to derive the balanced masses, and deviations from the means in these forecasts are proxies for background error (the Figure 2 caption gives details). Between the times 1230 and 1630 a cold front bringing precipitation passed through the domain. Note that, even though data from the UM are shown in Figure 2 (a, b), the Met Office procedure of accompanying the balance relation by vertical regression is not performed here.

In (a) the balanced mass variance (dashed black line) is consistently smaller than the total variance (black continuous line), apart from at the largest scales where the balanced mass dominates[3]. The unbalanced variance (dashed grey line) is mainly coincident with the total variance. The anomaly of the covariance model (continuous grey line) is of the same order as (but slightly larger than) the balanced

[3]The balanced mass at the largest scale is manually set to the level mean total mass. This is the undetermined integration constant in computing the balanced mass and means that the unbalanced variances are precisely zero at this scale.

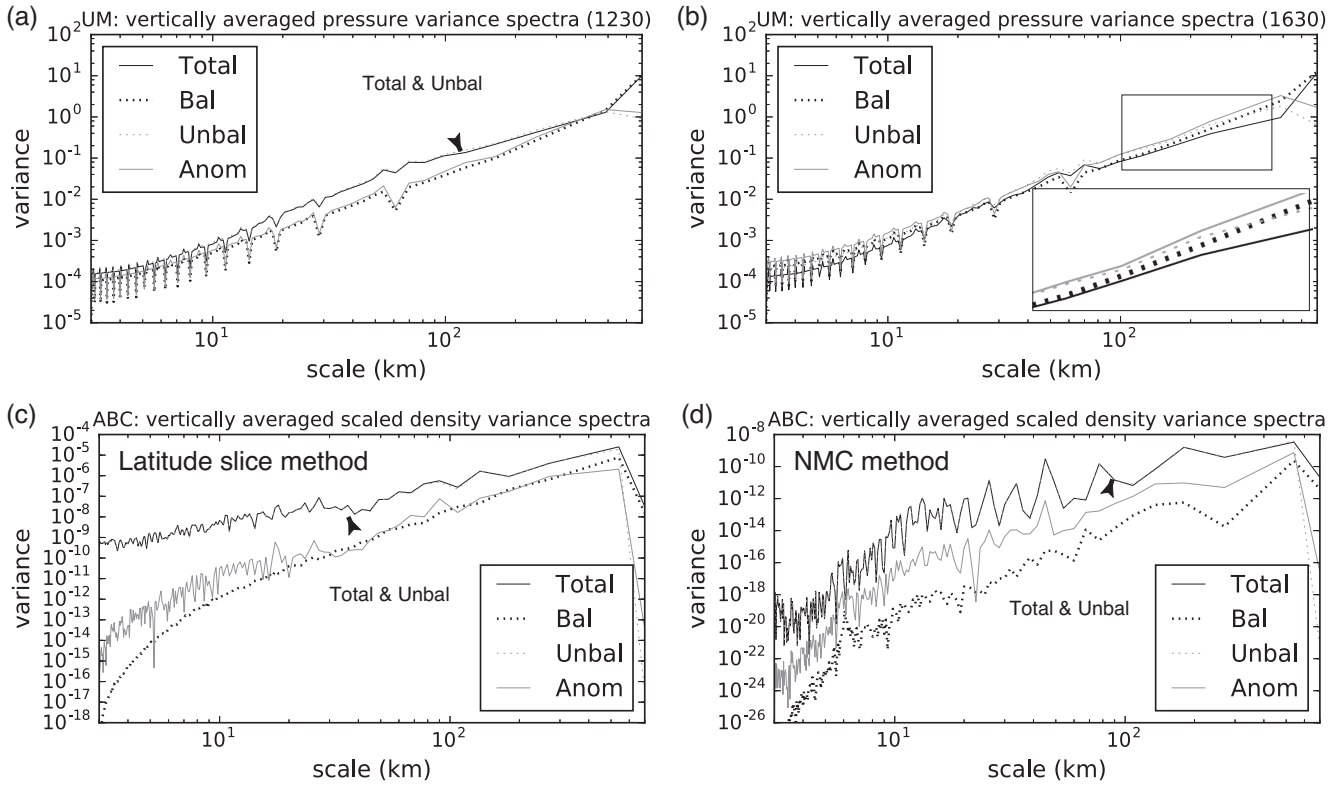

FIGURE 2 Total, balanced, and unbalanced contributions to the mass background-error variance as a function of scale in the UM at two different times (a) 1230 UTC and (b) 1630 UTC, and in the ABC model using two different forecast error samples (c) the latitude slice method and (d) the NMC method. The anomalous mass variance is also plotted (the magnitude of the difference between the total mass variance and the implied mass variance). The grid box sizes of the UM and ABC model are the same (1.5 km), although the UM's horizontal plane is two-dimensional (so the total horizontal wavenumber is $\sqrt{k_x^2 + k_y^2}$), and the ABC model's horizontal plane is one-dimensional. The downward spikes in the UM data are believed to be an artifact of the total wavenumber binning. The mass variables are pressure and scaled density in the UM and ABC respectively, and the variances are averaged over a few levels around 4 km in height. The UM spectra are found from 24-member ensembles of 1 hr UM forecasts initialised from an Ensemble Transform Kalman Filter (Baker *et al.*, 2014), where the balanced pressure is derived from the UM's horizontal winds (Appendix A1) by imposing Neuman boundary conditions. The ABC spectra are found from the two 260-member ensembles as detailed in Section 5 as used later to calibrate the DA system. The balanced scaled density is found from Equation (1). Note that, where the 'unbalanced' line is not visible, it overlaps with the 'total' line (as indicated), and enlarged parts of the spectra are shown in (b)

variance, except for the very largest scales. Where the anomaly is much smaller than the total (~ 20 to ~ 200 km), scenario (a) applies approximately in Figure 1. At smaller scales we suggest that scenario (b) applies, owing to the small balanced variance. In (b) there is a more varied ordering of the lines. Unlike in (a), the balanced variance does not take the lowest value, except marginally for scales between 40 and 80 km. However, at all other scales apart from the largest, the anomalous variance dominates. This suggests either scenarios (b) or (d) of Figure 1 are relevant for this panel.

In (c), where the latitude slice data are studied for the ABC model, the total and unbalanced variances dominate over all scales, apart from the largest scale, where only the total variance dominates, rather like the UM in (a). There appear to be three scale regimes. At the largest scales (above 300 km), the anomalous variance is the smallest. Between scales 30 and 300 km, the anomaly is of the same order of magnitude as the balanced variance, but at scales smaller than 30 km the balanced variance naturally takes smallest values, leaving the unbalanced variance to take all the variance. The small value of the anomaly at all scales suggest that scenario Figure 1a is valid, although for many scales it is because the balanced variance is naturally negligible anyway. In (d), where the NMC variances are shown, there are some qualitative similarities and differences with (c). As in (c), the total and unbalanced variances dominate and the balanced variance has the smallest contribution over all scales, apart from at the largest scales. In (d) the anomalous variance still contributes a significantly larger proportion than the balanced variances over a wide range of scales, but is still orders of magnitude smaller than

the total variance, hinting that the covariance model may still be appropriate. Note that, despite the very different scales of the y-axes between (c) and (d) when spectrally resolved, the standard deviations of the NMC $\widetilde{\rho}'$ errors are not smaller by orders of magnitude; the typical standard deviations of $\widetilde{\rho}'$ errors from the NMC method are about one third of those from the latitude slice method.

Where anomalous variances comprise a significant proportion of the total variance, this suggests that the LBE is not appropriate for the **B** model. These results show that, in terms of the validity of the LBE, there are some similarities and differences between the behaviours of a realistic 3D model like the UM and a 'toy' model like ABC, but that the applicability of the LBE in covariance modelling is likely to be less useful in a real system than in ABC. The rest of this paper is concerned only with ABC, but the above differences must be borne in mind when judging the relevance of our ABC DA results later in the paper to more realistic systems.

7 | DATA ASSIMILATION EXPERIMENTS WITH ABC WITH DIFFERENT BALANCE OPTIONS

In Section 6 we showed that the way that background errors are modelled with balance conditions may often be inappropriate and so requires further examination. In this section we describe and show results from a series of multi-cycle DA OSSEs (Observation System Simulation Experiments) with the ABC-DA system to see if linear and hydrostatic balances provide any advantage in covariance modelling at convective scales. This is done by switching on and off these balances in the **B**-matrix so that the total $\widetilde{\rho}'$ or b' increments are best represented either as the sum of balanced and unbalanced components as shown in Section 4 (e.g., $\alpha = 1$, $\beta = 1$), or each as just one component (the unbalanced component) describing the whole increment (e.g., $\alpha = 0$, $\beta = 0$). Turning off a balance relation may indeed allow a more realistic autocovariance for either $\widetilde{\rho}'$ or b', but this may come at the expense of destroying any multivariate components in the implied **B**-matrix. For each combination of α and β, and for each population of forecast errors described in Section 5, the spatial transforms mentioned in Section 4 are re-calibrated in the way described in Bannister (2020).

7.1 | Description of the experiments

We perform two sets of four DA experiments. The first set uses a the latitude slice method as a population of possible forecast errors to calibrate **B**, and the second set uses the NMC method. The four DA experiments in each set are as follows:

(a) $\alpha = 0$, $\beta = 0$ (no balance equations used),
(b) $\alpha = 0$, $\beta = 1$ (only HBE used),
(c) $\alpha = 1$, $\beta = 0$ (only LBE used), and
(d) $\alpha = 1$, $\beta = 1$ (LBE and HBE used).

The DA cycling period is 1 hr, where the analysis from each cycle is used to initialise a 1 hr forecast to yield the background of the next cycle, and 30 cycles are made. The first background is formed from the initial truth state with an initial perturbation of five times a random background error drawn from the modelled **B**-matrix. Observations of v, $\widetilde{\rho}'$, and b' are made at the start of each cycle, each on a 90×30 grid spanning the lower boundary to 12 km. The observation-error standard deviations for v, $\widetilde{\rho}'$, and b' are $0.5\,\mathrm{ms}^{-1}$, 10^{-3} and $1.5 \times 10^{-2}\mathrm{ms}^{-2}$, respectively.

7.2 | Triple observation assimilation increments

In order to help understand the impact of assimilating observations of v, $\widetilde{\rho}'$, and b' in each of the four DA experiments, Figure 3 shows the analysis increments of assimilating an individual triple observation of v, $\widetilde{\rho}'$, and b'. For this experiment, the **B**-matrix derived from the latitude slice method is used[4], and components of the innovation vector are set to the respective observation-error standard deviations.

When $\alpha = 0$ and $\beta = 0$ (first column), the increments are formed independently as the **B**-matrix in this case is univariate. The increments peak at the observation locations (crosses) and have negative side lobes. The different quantities have different length-scales, found from the calibration. When the HBE only is used to model the **B**-matrix ($\alpha = 0$ and $\beta = 1$, second column), the difference from the first column is mainly in the $\widetilde{\rho}'$ increment, but the b' increment is also slightly affected. This is due to the HBE coupling these two quantities as Equation (2). When the LBE only is used ($\alpha = 1$ and $\beta = 0$, third column) the differences from the first column are with the v and $\widetilde{\rho}'$ increments. These increments have lost their horizontal symmetry: the $\widetilde{\rho}'$ observation will encourage a dipole in v in order to maintain linear balance, but the v observation will encourage a positive v at the cross, thus shifting the dipole horizontally. When both HBE and LBE are used ($\alpha = 1$ and $\beta = 1$, fourth column) all fields are coupled and so all observations will affect all fields, giving rise again

[4]Results for the NMC-derived **B**-matrix show smaller magnitudes to those in Figure 3, but of similar structures.

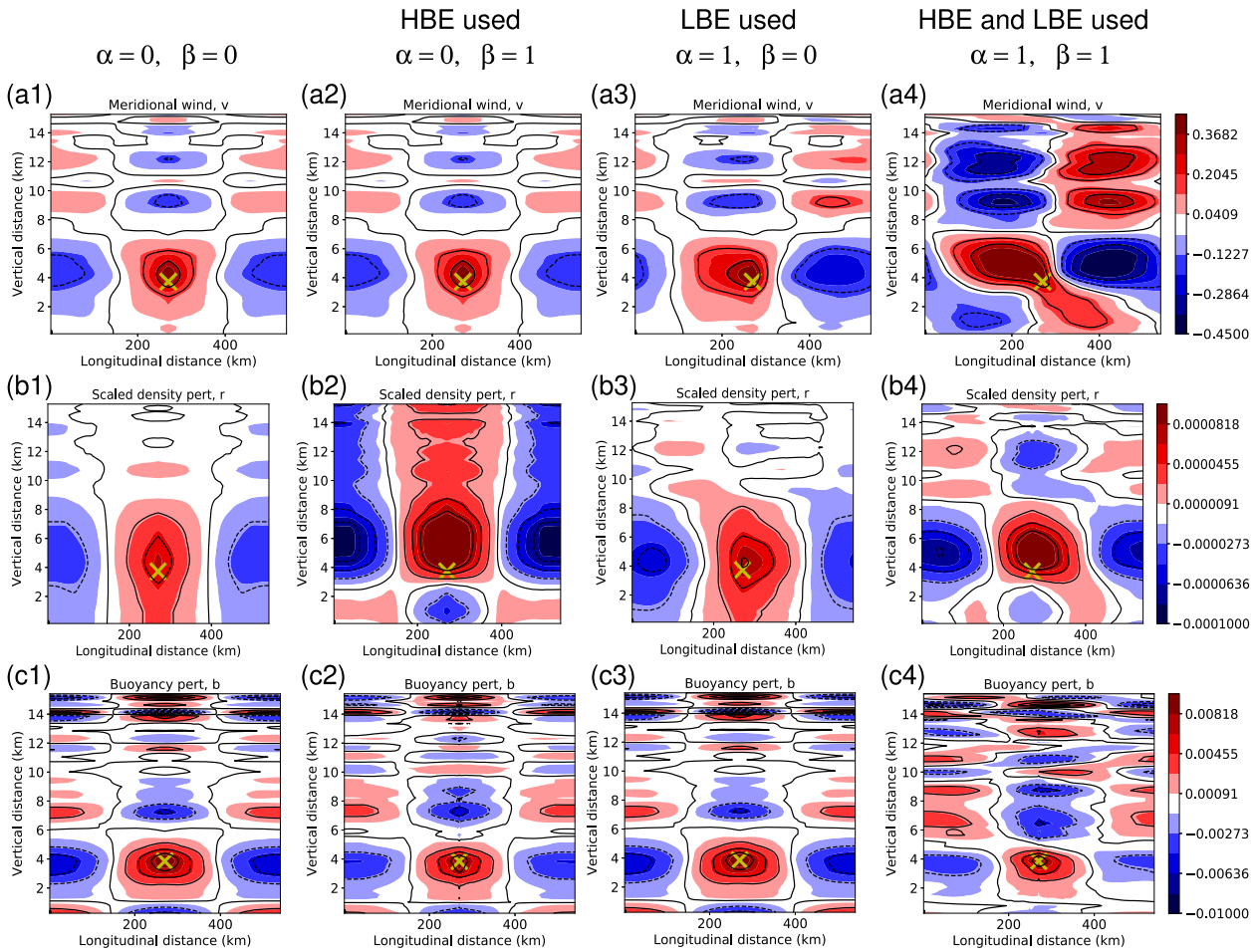

FIGURE 3 Analysis increments resulting from the assimilation of the triple observation (v, $\widetilde{\rho}'$, and b' combined) in the lower part of the domain (the position of the cross 4 km above the surface). The four columns correspond to the different **B**-matrices (configurations (i)-(iv) described in Section 7.1), which have different balances turned on/off, and each calibrated according to data from the latitude slice method. The rows show v, $\widetilde{\rho}'$ and b' respectively (other quantities have zero increments in these experiments) [Colour figure can be viewed at wileyonlinelibrary.com]

to the asymmetry. The question is whether imposing such structures has a beneficial or harmful effect in cycled DA for this system.

7.3 | ABC assimilation performance

7.3.1 | Variation with time of domain-averaged analysis errors

Figure 4 shows the domain-averaged root-mean-squared errors (RMSE) of each model quantity (rows) as a function of time for 0-to-1 hr forecasts starting from each cycled analysis (the times of the vertical yellow lines). The left column is for the systems calibrated with the latitude slice method and the right column for the NMC method. Recall that v, $\widetilde{\rho}'$, and b' are observed at the start of each cycle, but the unobserved quantities u and w are affected via the forecasts for each new background state. For conciseness, the experiments will be referred to as $[\alpha, \beta]$.

According to these statistics, there is no clear picture regarding which configuration of the **B**-matrix, $[\alpha, \beta]$, has lowest RMSE as the best and worst configurations differ between quantities and calibration methods. As the systems are quite complex, it is usually not possible to easily explain why some of the results appear, but there is still potentially useful information in just describing them.

- For errors in u: in (a1) $[0, 0]$ and $[1, 1]$ are the best settings, but in (a2) only $[1, 1]$ is the best. The fact that in (a1) turning on LB and HB separately worsens the RMSE, but turning them on together restores the performance must be due to the interaction between the two settings. This permits useful v–b' and $\widetilde{\rho}'$–b' covariances which enable extra information to be extracted from the observations, perhaps in some sense cancelling

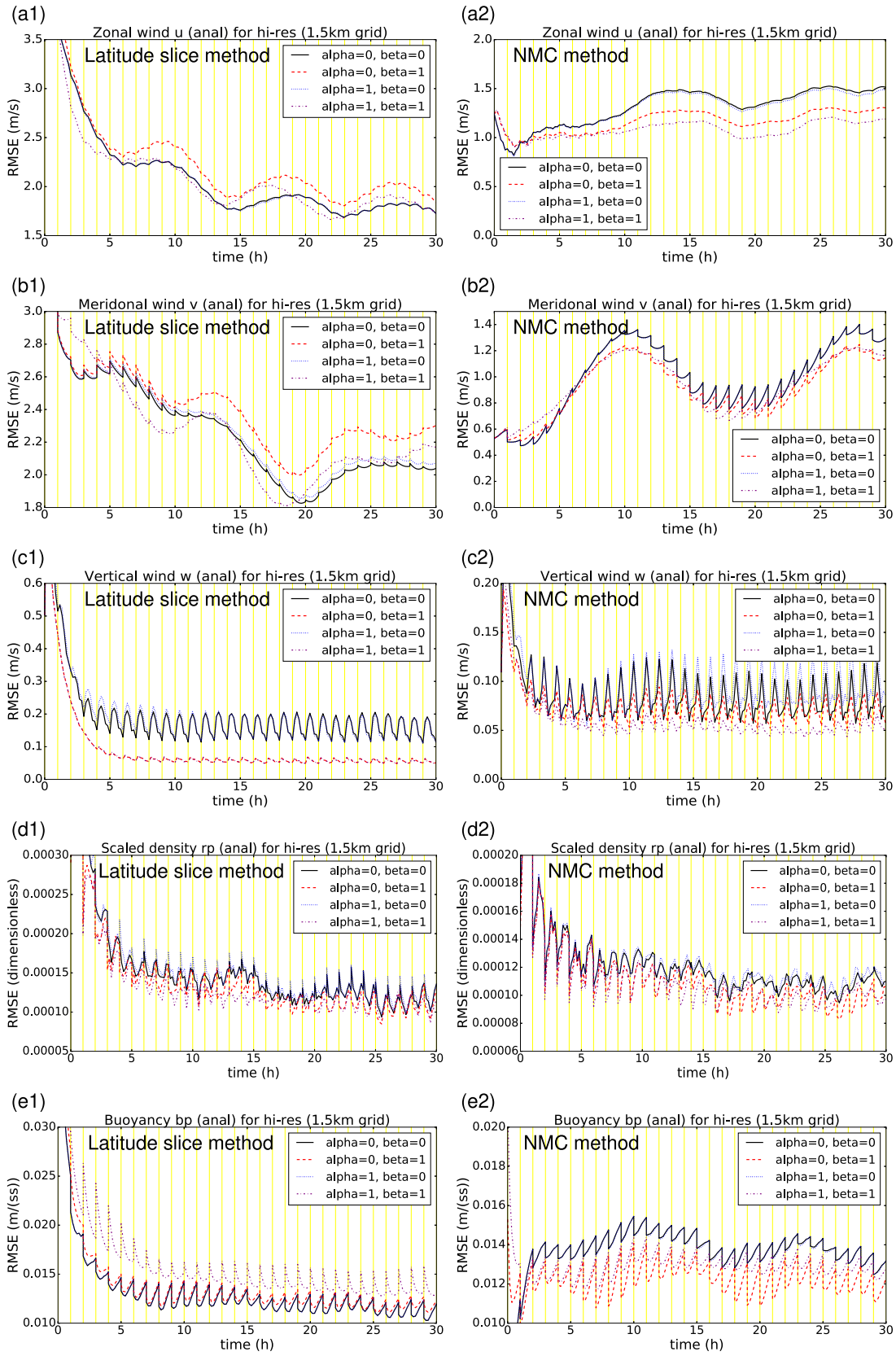

FIGURE 4 Cycled analysis errors for systems with **B**-matrices calibrated with the latitude slice method (left column) and the NMC method (right column) (Section 5). Different model quantities are in each row. Errors are defined as the domain-averaged root-mean-squared error (against the true state) of 0-to-1 hr forecasts starting from the analysis at each cycle. The yellow vertical lines mark the analysis times [Colour figure can be viewed at wileyonlinelibrary.com]

out the individual LB and HB related structures (see the full implied covariance matrix, equation (33) of Bannister (2020) with γ (a further switch) set to zero in that equation). Note that, since u is always univariate and unobserved, it is not updated by the DA, and so information is propagated from these quantities to u by the ABC model from one cycle to the next.

- For errors in v: in (b1) $[0,0]$ and $[1,1]$ are again the best settings, but in (b2) $[0,1]$ and $[1,1]$ are the best. A similar discussion to that for u applies, but here the covariances can affect v directly.
- For errors in w: in (c1) $[0,1]$ and $[1,1]$ are the best settings, but in (c2) only $[1,1]$ is the best. All of these settings invoke HB covariances. As for u, w is always univariate and unobserved and so is affected indirectly.
- For errors in $\widetilde{\rho}'$: in both (d1) and (d2) $[0,1]$ and $[1,1]$ are the best settings, that is, those that invoke HB covariances.
- For errors in b': in (e1) $[0,0]$, $[0,1]$, and $[1,0]$ are the best settings, but in (e2) only $[0,1]$ is the best. Applying only HB appears to be useful in all cases.

In general, the best compromise configuration is $[1,1]$ (the exception being the b' analysis, especially with the latitude slice calibration method). The same conclusions are reached by studying 1-to-2 hr forecast errors (instead of the 0-to-1 hr forecasts above; not shown).

7.3.2 | Spectrally resolved analysis-error variances

In order to shed some light on these results, Figure 5 shows snapshots ($t = 15$ hr) of error variance spectra for each quantity averaged over all vertical levels. The spectra are found by Fourier transforming the errors at the snapshot time to give $\epsilon(k,z)$ ($\epsilon \in \{\delta u, \delta v, \delta w, \delta\widetilde{\rho}', \delta b'\}$), where k is the wavenumber and z is the level, and then averaging $\left[\epsilon(k,z)\epsilon^*(k,z)\right]^{1/2}$ over z. The spectra are plotted as a function of wavelength, $2\pi/k$. This is done for each quantity, for each LB/HB switch configuration, and for the latitude slice (left column) and NMC (right column) methods of calibrating **B**.

- For spectral error variances in u: in (a1) $[0,0]$ and $[1,0]$ are the best settings, which are evident at small scales, but in (a2) $[1,1]$ is marginally the best setting, which is evident at large scales. Given the log scales, and the fact that the variances reduce with scale, any settings that show superiority only at small scales will not necessarily be seen in the domain averages in Figure 4. This is relevant when comparing Figure 5a1 with Figure 4a1. The result for (a2), showing superiority of $[1,1]$, *can* though be seen in Figure 4a2, since the difference is at large scales, which contribute most to the overall error variance. Figure 5a1 suggests that enforcing HB can be harmful if one is interested in small-scale analyses of u.
- For spectral error variances in v: in Figure 5b1 all settings appear to perform equivalently, but $[0,1]$ shows a slight increase in error at large scales. This is consistent with Figure 4b1 at $t = 15$ (computing similar spectra for $t = 18$ (not shown), reveals $[1,1]$ is the best setting at large scales, which is consistent with Figure 4b1 at that time). In Figure 5b2 there is a contrast between large and small scales – at large scales $[0,1]$ and $[1,1]$ are best (consistent with Figure 4b2), but at small scales $[0,0]$ and $[1,0]$ are best. The Figure 5b2 result suggests that enforcing HB can be helpful if one is interested in large-scale analyses of v, but harmful if one is interested in small-scale analyses.
- For spectral error variances in w: in Figure 5c1 there is also a contrast between large and small scales – at large scales $[0,1]$ and $[1,1]$ are the best (consistent with Figure 4c1), but at small scales $[0,0]$ and $[1,0]$ are best. In Figure 5c2 $[1,1]$ is best at large scales (consistent with Figure 4c2). In general these results again suggest a helpful role of HB in the covariances at large scales, but a harmful one at small scales. As w is not observed, these effects follow indirectly.
- For spectral error variances in $\widetilde{\rho}'$: in Figure 5d1, d2 the best setting is $[1,1]$ (for (d2) this is only marginally the best), which is consistent with Figure 4d1,d2.
- For spectral error variances in b': in Figure 5e1 the best settings are $[0,0]$ and $[1,0]$, which appear at small scales, and in (e2) the best settings are (marginally) $[0,1]$ and $[1,1]$ at large scales. Both of these results are consistent with Figure 4 e1,e2 at $t = 15$ hr. Once again, in general we find a helpful role of HB in the covariances at large scales, but a harmful one at small scales.

Although the performance of the different balance settings sometimes varies between the two calibration methods, there is a frequently observed characteristic that the HBE in particular remains a useful relationship to use above scales 10 to 100 km, and sometimes is harmful below these scales.

8 | CROSS-CORRELATIONS BETWEEN CONTROL PARAMETERS

In order to relate the results of Section 7 to the discussion in Sections 4 and 6, we briefly look at the correlations between the control parameters. Recall that anomalous

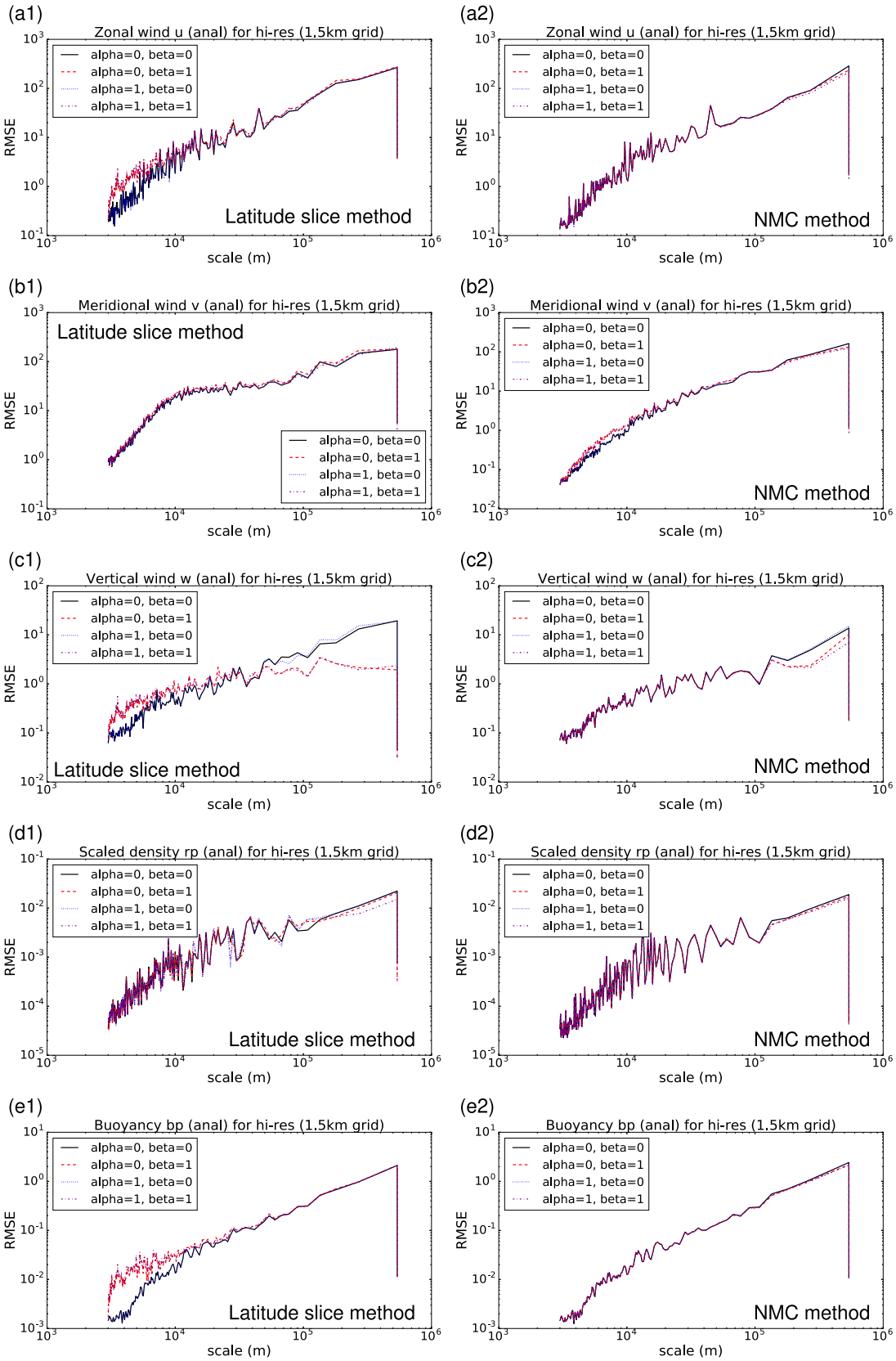

FIGURE 5 Spectra of analysis error variances for each model quantity (rows) at the snapshot mid-way through the experiment ($t = 15$ hr), averaged over all vertical levels. Results are shown for the systems with **B**-matrices calibrated with the latitude slice method (left column) and the NMC method (right column) (Section 5) [Colour figure can be viewed at wileyonlinelibrary.com]

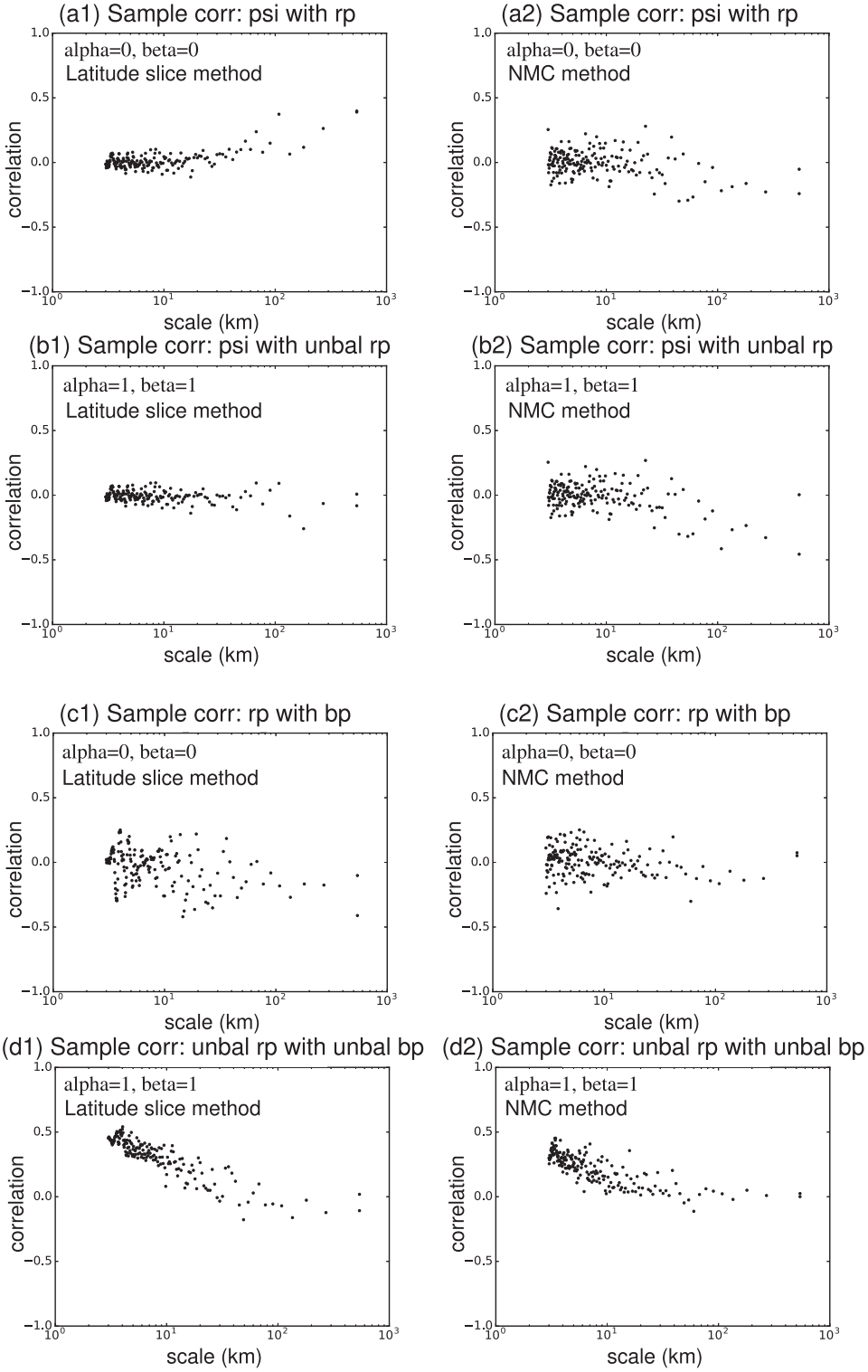

FIGURE 6
Cross-correlations of proxy 1 hr forecast errors of a selection of control parameters in spectral space. Proxies from the latitude slice method are shown in the left column and the NMC method in the right column (Section 5). The experiments are $\alpha = 0$, $\beta = 0$ (with control parameters $\delta\psi$ and $\delta\widetilde{\rho}'$ in (a1,a2); $\delta\widetilde{\rho}'$ and $\delta b'$ in (c1,c2)) and $\alpha = 1$, $\beta = 1$ (with control parameters $\delta\psi$ and $\delta\widetilde{\rho}'^{\mathrm{u}}$ in (b1,b2); $\delta\widetilde{\rho}'^{\mathrm{u}}$ and $\delta b'^{\mathrm{u}}$ in (d1,d2)). The correlations are computed for each level and then averaged vertically over all levels

(co)variances appear when there are correlations between the control variables. It was taken that assuming these correlations are zero when they are actually non-zero leads to sub-optimalities in the covariance model, possibly leading to enhancement of errors in the DA.

Figure 6 shows correlations (in spectral space) between a selection of control parameters calculated with each calibration method to show how these change with scale and with the different balance configurations used (for brevity, only [0, 0] and [1, 1] are considered). The first row is the correlation of $\delta\psi$ with $\delta\widetilde{\rho}'$ for [0, 0]. The correlations are close to zero for the latitude slice (a1) and NMC (a2) methods at scales smaller than 30 km, but deviate away from zero at larger scales (the NMC method in (a2)

shows a slight negative correlation trend, which is puzzling). The second row is the correlation of $\delta\psi$ with $\delta\widetilde{\rho}'^{\mathrm{u}}$ for [1, 1]. Clear from a comparison of the panels is that, whether or not the LBE is used, there is little effect on the correlations at small scales. Figure 6b1 shows that the large-scale non-zero correlations for [0, 0] are reduced. This is consistent with the lowering of the large-scale analysis errors of $\widetilde{\rho}'$ in Figure 5(d1) when the LBE is exploited (compare black and dotted purple lines). The NMC correlations are made very slightly more negative at large scales, even though there is also a reduction in analysis errors of $\widetilde{\rho}'$ in Figure 5d2 when the LBE is exploited.

The third row is the correlation of $\delta\widetilde{\rho}'$ with $\delta b'$ for [0, 0]. The correlations for the latitude slice (c1) and the NMC (c2) methods are similar and are distributed about zero. The fourth row is the correlation of $\delta\widetilde{\rho}'^{\mathrm{u}}$ with $\delta b'^{\mathrm{u}}$ for [1, 1]. Figure 6d1, d2 both show the emergence of significant correlations between these unbalanced parameters at small scales, and a tightening of correlations about zero at larger scales, which is consistent with an improvement of a number of quantities in Figure 5 (most notably Figure 5c1), but also in other panels where the dotted purple lines are lower than the black lines at small scales and/or the reverse at large scales).

9 | DISCUSSION, CONCLUSIONS AND FUTURE DIRECTIONS

9.1 | Discussion and conclusions

This paper explores some of the dynamical issues regarding background-error covariances in convective-scale DA. This has been done primarily using the simplified (longitude/height domain) ABC model and its 3D-Var DA system (known as ABC-DA; Bannister 2020). This paper is likely to be of interest to centres developing or operating convective-scale variational (3D or 4D) or hybrid DA systems. Particular questions relate to the use or not of balance equations to model background-error covariances, **B**.

We test whether separating fields into their balanced and unbalanced components (Section 3) can allow these motions to be treated separately in DA (as separate, uncorrelated, control parameters). Questions arise regarding their dynamical and statistical independence when the validity of the balance relationships is thought to be invalid, such as for linear balance (LB) and hydrostatic balance (HB) at km scales of motion. A conceptual framework is described to show how correlations that are present between such control parameters (but unaccounted for) will lead to anomalies in the background-error covariance model and to sub-optimalities in analyses (Sections 4 and 6). An analysis of some proxies of 1 hr forecast errors of the Met Office system and of the ABC system do suggest for instance that there could be significant anomalies in the background-error covariances in the UM and ABC systems, at least for the mass field.

A more definitive answer to whether the use of balance relationships at convective scale helps or hinders the DA problem is to perform cycled DA experiments (Section 7). A set of cycled ABC-DA identical-twin experiments have been performed on a midlatitude limited-area domain with a grid size of 1.5 km where LB and HB have been systematically switched off and on in **B** via the respective switches $\alpha = 0, 1$ and $\beta = 0, 1$, indicated by $[\alpha, \beta]$. Each experiment uses a separately calibrated **B**-matrix with training data from one of two methods: the 'latitude slice method' where proxies of forecast errors come from an ensemble of 1 hr forecasts whose members originate from latitude slices of a parent 3D model (a UM file), and the NMC method where proxies of forecast errors are differences between 2 hr and 1 hr ABC forecasts. The DA cycling time is 1 hr and observations of v, $\widetilde{\rho}'$ and b' are assimilated (analogous to observations of wind, pressure and temperature respectively).

Domain-averaged root-mean-squared errors indicate that there is no clear combination of LB and HB balance settings which gives the best results for all quantities. For instance [1, 1] performs better than [0, 0] for w and $\widetilde{\rho}'$ for both calibration techniques, but [1, 1] is worse than [0, 0] for b', but only for the latitude slice calibration method. The [1, 1] configuration is arguably the most successful overall for both calibration techniques. The positive effect of HB especially on w is particularly interesting as w is not observed and is a univariate variable in the DA. The positive effect is due to favourable influences of the assimilation with HB covariance structures on other variables and their coupling to w via the forecasts.

Looking at the contributions to errors at different scales reveals that using HB in particular in **B** is rarely a disadvantage at 'large scales' and not using balance is never a disadvantage at 'small scales'. The w errors in particular are interesting as, although HB provides an advantage on average, there is a clear disadvantage at 'small' scales (seen in the latitude slice calibrated results) and an advantage at 'large' scales (seen in both calibration techniques). The dividing line between 'large' and 'small' scales is found to be 10–100 km (depending on the quantity and calibration method). Scale-separated correlations between control parameters do show that HB does have uncorrelated control parameters at large scales, but univariate control variables (no balances enforced) lead to uncorrelated control parameters at small scales (Section 8). However, in the case of the LBE, setting $\alpha = 1$ naturally leads to a very small magnitude of $\delta\widetilde{\rho}'^{\mathrm{b}}$ at small scales anyway, and so here

it usually does not matter whether $\alpha = 0$ or 1 in the ABC system.

9.2 | Further analysis and possible future work

The results studied in this paper have focussed on switching on/off the LBE and HBE with α and β. Instead of treating α and β as switches (0 or 1) it is also possible to treat α and β as continuous variables, which can modulate the balanced components of the fields. In particular, one may choose 'optimal' values, which minimise the variance of the unbalanced variables and eliminate the cross-covariances between the balanced and unbalanced variables as found from training data (A.C. Lorenc, personal communication, 2021). These optimal values are

$$\alpha_{\text{opt}} = \left\langle \delta\widetilde{\rho}'^{\text{T}}(f/C)\delta\psi \right\rangle \Big/ \left\langle (f/C)\delta\psi^{\text{T}}(f/C)\delta\psi \right\rangle,$$
$$\text{and } \beta_{\text{opt}} = \left\langle \delta b'^{\text{T}}(-C\partial\delta\widetilde{\rho}'/\partial z) \right\rangle \Big/ \left\langle (-C\partial\delta\widetilde{\rho}'/\partial z)^{\text{T}}(-C\partial\delta\widetilde{\rho}'/\partial z) \right\rangle.$$

Here the angled brackets represent an average over the training data and recall that $(f/C)\delta\psi$ is the linearly balanced scaled density (Equation (1)), and $-C\partial\delta\widetilde{\rho}'/\partial z$ is the hydrostatically balanced buoyancy (Equation (2)). The α_{opt} factor is like a simplified version of the vertical regression operator which the Met Office applies with the linear balance operator (see text after Equation (4)). However, a rudimentary test of the configuration $[\alpha_{\text{opt}}, \beta_{\text{opt}}]$ in the data assimilation (involving a re-calibration) using both training ensembles – the latitude slice and NMC methods – did not result in a general decrease of RMSE in the analyses (not shown). Instead the $[\alpha_{\text{opt}}, \beta_{\text{opt}}]$ analyses were often found to be less accurate than the main settings [0, 0], [0, 1], [1, 0], and [1, 1].

This counterintuitive result may actually highlight a possibile difficulty with the training data rather than being due to a breakdown of the idea. This points to further work to generate better training ensembles, which more closely resemble forecast errors. Even though we know the 'truth' in our experiments, we do not know the true **B**-matrix, but it may be possible to approach it by bootstrapping with multiple iterations of the calibration. This presents an interesting possibility that, as the imposed **B**-matrix improves, the DA system may well become better at correcting one set of scales over another. For instance, if the large scales become better captured, then the analysis and forecast errors will become more unbalanced, which has reportedly been seen in NWP systems over recent decades (A.C. Lorenc, personal communication, 2021). We note that other major aspects, like the NWP models and the observing systems, have improved over that time as well as the ability to represent forecast-error covariances.

Returning to the interesting scale-dependence of the errors found in this paper's results, this suggests a multi-scale approach to the problem of background-error covariance modelling. One way forward is to use duplicate sets of control parameters – one set for small scales where no balance relations would be used (i.e., α, β are set to zero or small values), and another set for larger scales where the balance relations would remain as they are in current systems (since large-scale errors would still need to be corrected in a convective-scale system). A transition between the two could occur between 10 and 100 km wavelength, and the mathematical framework for this could be a two-band waveband transform (Fisher and Andersson, 2001; Deckmyn and Berre, 2005; Bannister, 2007; Pannekoucke *et al.*, 2007), with the smaller-scale band using univariate parameters (no balance relationships) and the larger-scale band using the balance relationships. These results may also be useful for ensemble-based convective-scale DA systems where efforts made in the field of **B** modelling as in this paper could be used as a multivariate and scale-dependent localisation scheme (Caron and Buehner, 2018).

Although we hope that this work is useful to the numerical weather prediction community, it has obvious limitations. Although the the ABC equations exhibit scale-dependent balance characteristics (Petrie *et al.*, 2017), it is a reduced-complexity, reduced-dimensionality model with dry dynamics. Incorporating moist processes into the model is work in progress and it will be interesting to see if similar results are obtained in such a revised model. The conclusions are also limited by the imperfect way that proxies of forecast errors (used to calibrate **B**) are generated. It will also be interesting to see how the results change with latitude, for example, approaching the Equator, where the unbalanced motions would be expected to have a larger effect.

ACKNOWLEDGEMENTS

This work has been supported by the U.K. Natural Environment Research Council via the National Centre for Earth Observation (grant LTS-S NE/R016518/1). The author would like to thank previous conversations with Mike Cullen, Mark Dixon, Stefano Migliorini, Alan O'Neill, and Ruth Petrie, and is grateful to the University of Reading Academic Computing Team for providing the computing resources. The author is also grateful for the comments of Andrew Lorenc and an anonymous referee.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest in connection with this work.

ORCID

Ross N. Bannister https://orcid.org/0000-0002-6846-8297

REFERENCES

Baker, L., Rudd, A., Migliorini, S. and Bannister, R.N. (2014) Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics*, 21, 19–39.

Ballard, S.P., Li, Z., Simonin, D. and Caron, J.-F. (2016) Performance of 4D-Var NWP-based nowcasting of precipitation at the Met Office for summer 2012. *Quarterly Journal of the Royal Meteorological Society*, 142, 472–487.

Bannister, R.N. (2007) Can wavelets improve the representation of forecast-error covariances in variational data assimilation?. *Monthly Weather Review*, 135, 387–408.

Bannister, R.N. (2008) A review of forecast-error covariance statistics in atmospheric variational data assimilation. II: modelling the forecast-error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134, 1971–1996.

Bannister, R.N. (2020) The ABC-DA system (v1.4): a variational data assimilation system for convective-scale assimilation research with a study of the impact of a balance constraint. *Geoscientific Model Development*, 13, 3789–3816.

Barker, D.M., Huang, W., Guo, Y.-R., Bourgeois, A. and Xiao, Q. (2004) A three-dimensional variational data assimilation system for MM5: implementation and initial results. *Monthly Weather Review*, 132, 897–914.

Berre, L. (2000) Estimation of synoptic and mesoscale forecast-error covariances in a limited-area model. *Monthly Weather Review*, 128, 644–667.

Berre, L., Ştefănescu, S.E. and Belo Pereira, M. (2006) The representation of the analysis effect in three error simulation techniques. *Tellus A*, 58, 196–209.

Bloom, S.C., Takacs, L.L., da Silva, A.M. and Ledvina, D. (1996) Data assimilation using incremental analysis updates. *Monthly Weather Review*, 124, 1256–1271.

Brousseau, P., Berre, L., Bouttier, F. and Desroziers, G. (2011) Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. *Quarterly Journal of the Royal Meteorological Society*, 137, 409–422.

Caron, J.-F. and Buehner, M. (2018) Scale-dependent background-error covariance localization: evaluation in a global deterministic weather forecasting system. *Monthly Weather Review*, 146, 1367–1381.

Caron, J.-F. and Fillion, L. (2010) An examination of background-error correlations between mass and rotational wind over precipitation regions. *Monthly Weather Review*, 138, 563–578.

Courtier, P. and Talagrand, O. (1990) Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus A*, 42, 531–549.

Cullen, M.J.P. and Davies, T. (1991) A conservative split-explicit integration scheme with fourth-order horizontal advection. *Quarterly Journal of the Royal Meteorological Society*, 117, 993–1002.

Deckmyn, A. and Berre, L. (2005) A wavelet approach to representing background-error covariances in a limited-area model. *Monthly Weather Review*, 133, 1279–1294.

Derber, J. and Bouttier, F. (1999) A reformulation of the background-error covariance in the ECMWF global data assimilation system. *Tellus A*, 51, 195–221.

Fisher, M. (2003). Background-error covariance modelling, *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean* 8–12 September 2003, pp. 45–63. Reading: ECMWF.

Fisher, M. and Andersson, E. (2001). Developments in 4D-Var and Kalman filtering. Technical memorandum 347, ECMWF, Reading, UK.

Gao, J., Xue, M., Shapiro, A. and Droegemeier, K.K. (1999) A variational method for the analysis of three-dimensional wind fields from two Doppler radars. *Monthly Weather Review*, 127, 2128–2142.

Gauthier, P., Buehner, M. and Fillion, L. (1999). Background-error statistics modelling in a 3D variational data assimilation scheme: estimation and impact on the analyses, *Proceedings of Workshop on Diagnosis of Data Assimilation Systems* 2–4 November 1998, pp. 131–145. Reading: ECMWF.

Ge, G., Gao, J. and Xue, M. (2012) Diagnostic pressure equation as a weak constraint in a storm-scale three-dimensional variational radar data assimilation system. *Journal of Atmospheric and Oceanic Technology*, 29, 1075–1092.

Heng, B.P., Tubbs, R., Huang, X.-Y., Macpherson, B., Barker, D.M., Boyd, D.F., Kelly, G., North, R., Stewart, L., Webster, S. and Wlasak, M. (2020) SINGV-DA: a data assimilation system for convective-scale numerical weather prediction over Singapore. *Quarterly Journal of the Royal Meteorological Society*, 146, 1923–1938. https://doi.org/10.1002/qj.3774.

Hodyss, D. and Nichols, N. (2015) The error of representation: basic understanding. *Tellus A: Dynamic Meteorology and Oceanography*, 67. https://doi.org/10.3402/tellusa.v67.24822.

Honda, Y., Nishijima, M., Koizumi, K., Ohta, Y., Tamiya, K., Kawabata, T. and Tsuyuki, T. (2005) A pre-operational variational data assimilation system for a non-hydrostatic model at the Japan Meteorological Agency: formulation and preliminary results. *Quarterly Journal of the Royal Meteorological Society*, 131, 3465–3475.

Houtekamer, P., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H.L. (1996) A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124, 1225–1242.

Hu, M., Xue, M. and Brewster, K. (2006) 3DVar and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: cloud analysis and its impact. *Monthly Weather Review*, 134, 675–698.

Ingleby, N.B. (2001) The statistical structure of forecast errors and its representation in the Met Office global 3D variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 127, 209–231.

Kleist, D.T., Parrish, D.F., Derber, J.C., Treadon, R., Errico, R.M. and Yang, R. (2009) Improving incremental balance in the GSI 3DVar analysis system. *Monthly Weather Review*, 137, 1046–1060.

Lorenc, A.C. (1981) A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109, 701–721.

Lorenc, A.C., Ballard, S.P., Bell, R., Ingleby, N.B., Andrews, P., Barker, D.M., Bray, J., Clayton, A., Dalby, T., Li, D., Payne, T. and

Saunders, F. (2000) The Met Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126, 2991–3012.

Lynch, P. and Huang, X.-Y. (1992) Initialization of the HIRLAM model using a digital filter. *Monthly Weather Review*, 120, 1019–1034.

Pannekoucke, O., Berre, L. and Desroziers, G. (2007) Filtering properties of wavelets for local background-error correlations. *Quarterly Journal of the Royal Meteorological Society*, 133, 363–379.

Parrish, D.F. and Derber, J.C. (1992) The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120, 1747–1763.

Petrie, R.E., Bannister, R.N. and Cullen, M.J.P. (2017) The 'ABC model': a non-hydrostatic toy model for use in convective-scale data assimilation investigations. *Geoscientific Model Development*, 10, 4419–4441.

Potvin, C.K. and Wicker, L.J. (2013) Correcting fast-mode pressure errors in storm-scale ensemble Kalman filter analyses. *Advances in Meteorology*, 2013. https://doi.org/10.1155/2013/624931.

Shen, F., Xu, D. and Min, J. (2019) Effect of momentum control variables on assimilating radar observations for the analysis and forecast for typhoon *Chanthu* (2010). *Atmospheric Research*, 230, 104622.

Song, H.-J. and Kang, J.-H. (2019) Effects of the wind–mass balance constraint on ensemble forecasts in the hybrid-4DEnVar. *Quarterly Journal of the Royal Meteorological Society*, 145, 434–449.

Sun, J. (2005) Convective-scale assimilation of radar data: progress and challenges. *Quarterly Journal of the Royal Meteorological Society*, 131, 3439–3463.

Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., Joe, P., Barker, D.M., Li, P.-W., Golding, B., Xu, M. and Pinto, J. (2014) Use of NWP for nowcasting convective precipitation: recent progress and challenges. *Bulletin of the American Meteorological Society*, 95, 409–426.

Thiruvengadam, P., Indu, J. and Ghosh, S. (2019) Assimilation of Doppler weather radar data with a regional WRF-3DVar system: impact of control variables on forecasts of a heavy rainfall case. *Advances in Water Resources*, 126, 24–39.

Tong, W., Li, G., Sun, J., Tang, X. and Zhang, Y. (2016) Design strategies of an hourly update 3DVar data assimilation system for improved convective forecasting. *Weather and Forecasting*, 31, 1673–1695.

Vetra-Carvalho, S., Dixon, M., Migliorini, S., Nichols, N.K. and Ballard, S.P. (2012) Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, 138, 1709–1720.

Wang, C., Chen, Y., Chen, M. and Shen, J. (2020) Data assimilation of a dense wind profiler network and its impact on convective forecasting. *Atmospheric Research*, 238. https://doi.org/10.1016/j.atmosres.2020.104880.

Xie, Y. and MacDonald, A.E. (2012) Selection of momentum variables for a three-dimensional variational analysis. *Pure and Applied Geophysics*, 169, 335–351.

Xu, D., Shen, F. and Min, J. (2020) Effect of background error tuning on assimilating radar radial velocity observations for the forecast of hurricane tracks and intensities. *Meteorological Applications*, 27(1). https://doi.org/10.1002/met.1820.

Yano, J.-I., Ziemiański, M.Z., Cullen, M.J.P., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S.L., Homar, V., K˝ ohler, M., Krichak, S., Michaelides, S., Phillips, V.T.J., Soares, P.M.M. and Wyszogrodzki, A.A. (2018) Scientific challenges of convective-scale numerical weather prediction. *Bulletin of the American Meteorological Society*, 99, 699–710.

Yu, X. and Lee, T.-Y. (2010) Role of convective parameterization in simulations of a convection band at grey-zone resolutions. *Tellus A: Dynamic Meteorology and Oceanography*, 62, 617–632.

Zakeri, Z., Azadi, M. and Ghader, S. (2018) The impact of different background errors in the assimilation of satellite radiances and in-situ observational data using WRFDA for three rainfall events over Iran. *Advances in Space Research*, 61, 433–447.

How to cite this article: Bannister, R.N. (2021) Balance conditions in variational data assimilation for a high-resolution forecast model. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2917–2934. Available from: https://doi.org/10.1002/qj.4106

APPENDIX A. SUMMARY OF BALANCE RELATIONSHIPS FOR EULER'S EQUATIONS

For reference purposes, this Appendix gives the balance equations for Euler's equations.

A.1 Linear balance in Euler's equations

The linear balance equation (LBE) derived from Euler's equations relates the balanced pressure (p^{b}), the horizontal wind ($\mathbf{u}_{\mathrm{h}}$), and the balanced specific mass (α^{b}):

$$\nabla_{\mathrm{h}}^2 p^{\mathrm{b}} - \mathbf{k} \cdot \nabla \times \left(f \mathbf{u}_{\mathrm{h}} / \alpha^{\mathrm{b}} \right) = 0, \tag{A1}$$

where ∇_{h} and ∇ are the horizontal and three-dimensional gradient operators respectively, f is the Coriolis parameter, $\mathbf{k}$ is the vertical unit vector, and $\mathbf{u}_{\mathrm{h}} = (u, v, 0)$. It is assumed that $\mathbf{u}_{\mathrm{h}}$ is completely balanced and so does not need a 'b' superscript. Equation A1 is equivalent to geostrophic balance when f and α^{b} are constant. For incremental DA the linearised form of Equation (A1) is needed:

$$\nabla_{\mathrm{h}}^2 \delta p^{\mathrm{b}} + \mathbf{k} \cdot \nabla \times \left\{ \left(f / {\alpha^{\mathrm{b}}}^2 \right) \mathbf{u}_{\mathrm{h}} \delta \alpha^{\mathrm{b}} \right\} - \mathbf{k} \cdot \nabla \times \left\{ \left(f / \alpha^{\mathrm{b}} \right) \delta \mathbf{u}_{\mathrm{h}} \right\} = 0, \tag{A2}$$

where the δ-prefixed variables are increments, and other variables comprise the linearisation state, for example, $\alpha^{\mathrm{b}} = \alpha^{\mathrm{b}} + \delta\alpha^{\mathrm{b}}$. The Met Office currently uses Equation (A2) with the assumption that $\delta\alpha^{\mathrm{b}}$ has a negligible contribution.

A.2 Hydrostatic balance in Euler's equations

The HBE, $\partial p/\partial z = -\rho g$, together with the equation of state

$$p = R_{\mathrm{d}}\rho\theta_{\mathrm{v}}(p/p_{1000})^{R_{\mathrm{d}}/c_{\mathrm{p}}},$$

may be written in a form relating balanced pressure (p^{b}) and balanced virtual potential temperature ($\theta_{\mathrm{v}}^{\mathrm{b}}$):

$$\frac{\partial}{\partial z}\left(\frac{p^{\mathrm{b}}}{p_{1000}}\right)^{\frac{R_{\mathrm{d}}}{c_p}} + \frac{g}{c_p\theta_{\mathrm{v}}^{\mathrm{b}}} = 0, \tag{A3}$$

where z is height, $p_{1000} = 1000\,\mathrm{hPa}$, R_{d} is the gas constant for dry air, c_p is the constant pressure specific heat capacity, and g is the acceleration due to gravity. The incremental form of Equation (A3) is:

$$\frac{R_{\mathrm{d}}}{c_p}\frac{\partial}{\partial z}\left[\left(\frac{p^{\mathrm{b}}}{p_{1000}}\right)^{\frac{R_{\mathrm{d}}}{c_p}}\frac{\delta p^{\mathrm{b}}}{p^{\mathrm{b}}}\right] - \frac{g}{{\theta_{\mathrm{v}}^{\mathrm{b}}}^2}\delta\theta_{\mathrm{v}}^{\mathrm{b}} = 0. \tag{A4}$$

This is the form used in the Met Office's variational scheme to diagnose temperature from pressure. Further, the Met Office's scheme assumes that all increments are exactly hydrostatically balanced, that is, $\delta p = \delta p^{\mathrm{b}}$ and $\delta\theta_{\mathrm{v}} = \delta\theta_{\mathrm{v}}^{\mathrm{b}}$.